Математическая модель оценки реактивности распределенной системы обработки информации с использованием метода декомпозиционной аппроксимации

А. Н. Скоба, В. К. Михайлов, С.А. Назаров, Н.С. Скорик

Южно-Российский государственный политехнический университет (НПИ) им. М. И. Платова, Новочеркасск

Аннотация: в данной статье с использованием метода декомпозиционной аппроксимации на основе теореме Нортона приведена, разработанная авторами математическая модель функционирования распределенной системы обработки информации (СОИ) на базе локальной вычислительной сети (ЛВС) файл-серверной архитектуры с произвольными функциями распределения времени обслуживания заявок в узлах сети. Представлена также разработанная концептуальная модель эквивалентной двухузловой сети массового обслуживания (СеМО) и основные математические отношения для вычисления интенсивности обслуживания в композиционном центре, а также выражения для вычисления среднего времени реакции системы на запросы пользователей.

Ключевые слова: распределенная система обработки информации, декомпозиционная аппроксимация, сеть массового обслуживания, композиционный центр, среднее время реакции системы, концептуальная модель, нормализующая константа.

Одним из основных недостатков применения точных методов для разработки математических моделей функционирования распределенных систем обработки информации (СОИ) является предположение об экспоненциальном законе распределения времени обслуживания заявок в узлах сети. Данное предположение требует проведения достаточной серьезной системы доказательств [1], что для реальных распределенных СОИ является достаточно трудоемкой задачей. Поэтому одним из альтернативных подходов для конструирования математических моделей функционирования распределенных СОИ с использованием аппарата сетей массового обслуживания, позволяющим существенно «ослабить» обязательное требование об экспоненциальности функции распределения, является применение приближенных методов. В их основе лежит либо аппроксимация произвольных законов распределения длительности обслуживания заявок в узлах сети обобщенным распределением Кокса [2], либо диффузионная или декомпозиционная аппроксимация [3,4], либо имитационное моделирование [5,6].

Как было отмечено в работе [7] метод анализа СеМО, основанный на аппроксимации функционий распределения длительности обслуживания заявок в центрах сети обобщенным распределением Кокса приводит к значительному увеличению мощности пространства состояний СеМО и, следовательно, может быть эффективно использован лишь для сетей малой размерности.

Методы имитационного моделирования ориентированы на воспроизведение алгоритма функционирования распределенной СОИ и позволяют описывать работу практически любых по степени сложности систем. Однако эти методы требуют большого количества реализаций для получения вероятности редких событий и соответственно значительных затрат времени. Так, согласно [5], для нахождения вероятности $p = 10^{-3}$ при заданной точности $\varepsilon = 10^{-4}$, необходимо провести около 384000 реализаций. В тоже время, для вычисления среднего значения случайной величины не требуется такая высокая точность, и, следовательно, необходимо выполнить существенно меньшее количество реализаций.

Сущность метода диффузионной аппроксимации [3] состоит в аппроксимации дискретного случайного процесса, описывающего число соответствий в центрах СеМО, непрерывном диффузионном процессом. Однако, при вычислении интегральных характеристик вычислительных систем со сравнительно небольшой загрузкой (что характерно для систем, работающих в реальном масштабе времени) метод дает недопустимую погрешность, тем большую, чем меньше загрузка центров СеМО.

Еще одним приближенным методом конструирования математических моделей функционирования распределенных СОИ, позволяющим преодолеть некоторые недостатки диффузионных моделей, является декомпозиционная аппроксимация, в основе которой лежит подход, основанный на теореме Нортона [8]. В рамках данного подхода приближенное исследование работы СеМО с произвольным законом распределения длительности обслуживания заявок в узлах сети осуществляется путем перехода к эквивалентной сети с двумя центрами обслуживания. При этом первый центр двухузловой сети совпадает с *i*-ым центром исходной сети $(i=\overline{1,M})$, а второй (композиционный), являющийся эквивалентом остающейся части сети, обладает экспоненциально распределенным временем обслуживания заявок с параметром $\mu_B(\overline{n}), \overline{n}=(n_1,...,n_M)$, зависящим от числа сообщений в нем.

Применим идею декомпозиционного подхода, изложенного в работе [?], для получения вероятностных характеристик распределенной СОИ, реализованной на базе архитектуры «файл-сервер» [9].

Концептуальная модель функционирования рассматриваемой распределенной СОИ представляет замкнутую СеМО, включающую: n_R пользователей, формирующих запросы с интенсивностями $\lambda_i(i=\overline{1,n_R})$; $M=2n_R+1$ обслуживающих центров, моделирующих работу пользователей, канала передачи данных и обслуживающих ПЭВМ.

Предполагается также, что рассматриваемая СеМО удовлетворяет условиям локального баланса с вероятностями перехода, задаваемыми матрицами переходных вероятностей $\|P_{ik}(r)\|$, $(i,k=\overline{1,2n_R+1}, r=\overline{1,n_R})$ и интенсивностями обслуживания в узлах сети $\mu_{sr}(s=\overline{1,2n_R+1}, r=\overline{1,n_R})$.

При расчете СеМО с помощью декомпозиционного подхода, основанного на Теореме Нортона [3], концептуальная модель исходной сети, заменяется эквивалентной сетью, представленной на рисунке 1.

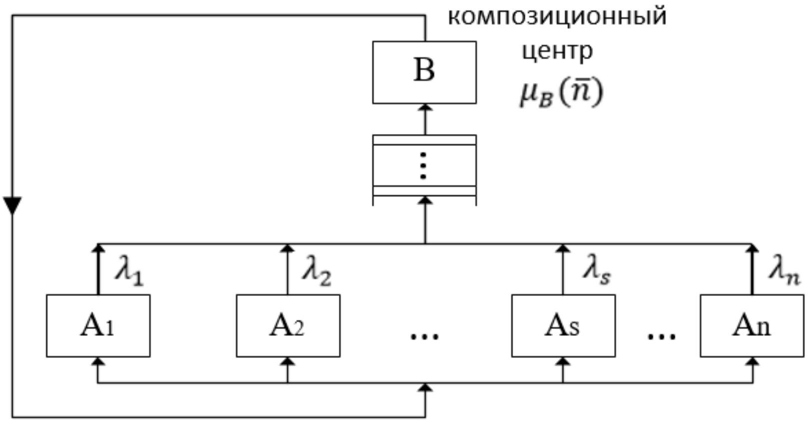

Рисунок 1 – Эквивалентная сеть с композиционным центром В

Дальнейший расчет осуществляется в следующей последовательности:

Шаг 1. Интенсивность обслуживания в дополнительном центре В вычисляется по формуле:

$$\mu_B(\overline{n_R}) = \sum_{r=1}^{n_R} e_{1r} \frac{g(\overline{n_R} - \overline{1_r}, m)}{g(\overline{n_R}, m)},$$

где $\overline{n_R} = (n_1, \ldots, n_R), n_i = \overline{0,1}$; $\overline{1_r} = (0,0,\ldots,1,\ldots,0,0)$ – вектор размерности n_R в r-ой координате которого стоит 1, а все остальные значения равны 0, $m = \overline{1, 2n_R + 1}$; величины $e_{sr}(s = \overline{1, 2n_R + 1}, r = \overline{1, n_R})$, находится из решения систем линейных алгебраических уравнений:

$$e_{sr} = \sum_{j=1}^{2n_R+1} e_{jr} P_{js}(r), s = \overline{1, 2n_R + 1}, r = \overline{1, n_R}.$$

Кроме того, при вычислении величин $g(\overline{n_R}, m)$, где $\overline{n_R} = (n_1, \ldots, n_R)$, $n_i = \overline{0,1}, m = \overline{1, 2n_R + 1}$, предполагается также, что $g(\overline{n_R}, m) = 1$, если все $n_i = 0 \quad (i = \overline{1, n_R})$, и $g(\overline{n_R} - \overline{1_r}, m) = 0$, если хотя-бы одна из координат вектора $\overline{n_R} - \overline{1_r} < 0$.

Шаг 2. Для вычисления нормализующей константы двухузловой сети воспользуемся соотношениями [7]:

$$g(\overline{n_R}, m) = \sum_{\overline{k_R}=0}^{\overline{n_R}} Z_m(\overline{k_R}) g(\overline{n_R} - \overline{k_R}, m-1)$$

или в развернутом виде:

$$g(n_1, n_2, ..., n_R; m) = \sum_{k_R=0}^{n_R} ... \sum_{k_1=0}^{n_1} Z_m(k_1, k_2, ..., k_R) \cdot g(n_1 - k_1, ..., n_R - k_R; m-1).$$

Для независящих от нагрузки центров:

$$Z_m(\overline{n_R}) = \sum_{r=1}^{R} x_{mr} Z_m(\overline{n_R} - \overline{1_r}),$$

где $x_{mr} = \dfrac{e_{sr}}{\mu_{sr}} (s = \overline{1, 2n_R + 1}, r = \overline{1, n_R}), g(\overline{n_R}, 1) = Z_1(\overline{n_R})$, а для зависящих от нагрузки центров:

$$Z_m(\overline{n_R}) = \left[\prod_{n_r=0}^{n_R} \frac{1}{\mu_m(\overline{n_R})}\right] n_R! \prod_{r=1}^{R} \frac{e_{mr}{}^{n_r}}{n_r!}.$$

В нашем случае: $n_i = \overline{0,1}, i = \overline{1,R}, m = B = 2$ и формулы принимают следующий вид:

$$g(n_1, n_2, ..., n_R; 2) = \sum_{k_R=0}^{1} ... \sum_{k_1=0}^{1} Z_2(k_1, k_2, ..., k_R) \cdot Z_1(n_1 - k_1, n_2 - k_2, ..., n_R - k_R),$$

а так как предполагается, что первый центр не зависит от нагрузки, то для него:

$$Z_1(\overline{n_R}) = \sum_{r=1}^{R} x_{1r} Z_1(\overline{n_R} - \overline{1_r}),$$

а второй центр (центр B) зависит от нагрузки, и для него:

$$Z_2(\overline{n_R}) = \left[\prod_{n_r=0}^{1} \frac{1}{\mu_B(\overline{n_R})}\right] n_R! \prod_{r=1}^{R} e_{2r}{}^{n_r},$$

где $\overline{n_R} = (n_1, ..., n_R), n_R! = (n_1 + ... + n_R)!, n_r = \overline{0,1}$.

Согласно [9] выражение для расчета среднего времени реакции системы на запросы пользователей $\overline{T}$ может быть приведено к виду:

$$\overline{T} = \left(\frac{1}{\sum_{r=1}^{n_R} \lambda_r}\right) \sum_{r=1}^{n_R} \frac{1 - P_r(1)}{P_r(1)},$$

где $P_r(1) = \sum_{n_1=0}^{1} ... \sum_{n_{r-1}=0}^{1} \sum_{n_{r+1}=0}^{1} \sum_{n_R=0}^{1} P_1(\overline{i_1}^{(r)}(1), \overline{n_R})$, $\overline{i_1}^{(r)} = (n_1, ..., n_{r-1}, 1, n_{r+1}, ..., n_R)$. В работе [9] было показано, что выражение для $P_1(\overline{i_1}^{(r)}(1), \overline{n_R})$ может быть приведено к виду:

$$P_1(\overline{i_1}^{(r)}(1),\overline{n_R})=\frac{Z_1(\overline{i_1}^{(r)}(1),\overline{n_R})}{g(\overline{n_R},B)}\left[g(\overline{n_R}-\overline{i_1}^{(r)}(1),B)-\sum_{r=1}^{n_R}x_{1r}\,g(\overline{n_R}-\overline{1_r}-\overline{i_1}^{(r)}(1),B)\right], \quad (1)$$

где $x_{1r}=\dfrac{e_{1r}}{\mu_{1r}}$.

Как видно из (1), расчет величины $\overline{T}$ сводится, по существу, к расчету нормализующей константы $g(\overline{n_R},B)$ для сети с существенно меньшим пространством состояний, чем это имело место для модели, представленной в работе [9], для вычисления которой может быть использован рекуррентный метод Бузена [10].

Литература

1. Зуев В.А., Панфилов А.Н., Скоба А.Н. Методика статистического анализа характеристик входных потоков запросов в системах обработки информации // Инженерный вестник Дона. 2015. №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2900.

2. Бусленко Н.П., Калашников В.В., Коваленко И.Н. Лекции по теории сложных систем. – М.: Сов. Радио, 1973. – 440 с.

3. Авен О.И., Гурин Н.И., Котан Я.А. Оценка качества и оптимизации вычислительных систем. – М.: Наука, 1982. – 464 с.

4. Крутликов В.К., Тарасов В.Н. Анализ и расчет сетей массового обслуживания с использованием двумерной диффузионной аппроксимации // Автоматика и телемеханика, 1983. – №8 – с. 72-84.

5. Манусевич В.С., Бусленко Н.П. Имитационное моделирование сетей массового обслуживания // Методы развития теории телетрафика. – М.: Наука, 1979. – с. 8-18.

6. Sauer C.H., Machair E.A., Hurouse J.F. Queueing Network Simulation of Computer Communication // IEEE J. Selected Areas in Commun, 1984. – V. SAC-2, N 1. – pp. 203-219.

7. Жожикашвили В.А., Вишневский В.М. Сети массового обслуживания. Теория и применение к сетям ЭВМ. – М.: Радио и связь, 1988. – 192 с.

8. Вишневский В.М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512 с.

9. Скоба А.Н., Состина Е.В. Математическая модель оптимального размещения распределённой базы данных по узлам ЛВС на базе файл-серверной архитектуры // Инженерный вестник Дона. 2015. №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2881.

10. Buzen J.P. Computational Algorithms for Closed Queueing Networks with Exponential Servers. Commun. ACM. 1983. Vol.16, №9. – pp.527-531.

References

1. Zuev V.A., Panfilov A.N., Scoba A.N. Inženernyj vestnik Dona (Rus), 2015, №2 URL: ivdon.ru/ru/magazine/archive/n2y2015/2900.

2. Buslenko N.P., Kalashnikov V.V., Kovalenko I.N. Lekcii po teorii slozhny`x sistem [Lectures on the theory of complex systems]. M.: Sov. Radio, 1973. 440p.

3. Aven O.I., Gurin N.I., Kotan Ya.A. Ocenka kachestva i optimizacii vy`chislitel`ny`x sistem [Evaluation of the quality and optimization of computing systems]. M.: Nauka, 1982. 464p.

4. Krutlikov V.K., Tarasov V.N. Avtomatika i telemexanika, 1983. №8. pp. 72-84.

5. Manusevich V.S., Buslenko N.P. Imitacionnoe modelirovanie setej massovogo obsluzhivaniya [Simulation modeling of Queuing networks]. Metody` razvitiya teorii teletrafika. M.: Nauka, 1979. pp. 8-18.

6. Sauer C.H., Machair E.A., Hurouse J.F. IEEE J. Selected Areas in Commun, 1984. V. SAC-2, N 1. pp. 203-219.

7. Zhozhikashvili V.A., Vishnevskij V.M. Seti massovogo obsluzhivaniya. Teoriya i primenenie k setyam E`VM [Queueing networks. Theory and its network application]. M.: Radio i svyaz', 1988. 192p.

8. Vishnevskij V.M. Teoreticheskie osnovy` proektirovaniya komp`yuterny`x setej [Theoretical foundations of computer network design]. M.: Texnosfera, 2003. 512p.

9. Skoba A.N., Sostina E.V. Inženernyj vestnik Dona (Rus), 2015, №2. URL:ivdon.ru/ru/ magazine/archive/n2y2015/2881.

10. Buzen J.P. Computational Algorithms for Closed Queueing Networks with Exponential Servers. Commun. ACM. 1983. Vol.16, №9. pp. 527-531.